41
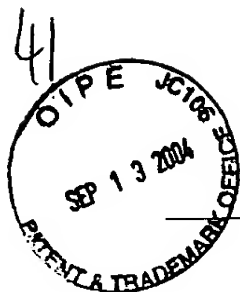
AF/3655

PATENT

Case Docket No. AMAZON.059A
Date: September 10, 2004

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Nicholas Lee
Appl. No.	:	09/650,173
Filed	:	August 29, 2000
For	:	VOICE INTERFACE AND METHODS FOR IMPROVING RECOGNITION ACCURACY OF VOICE SEARCH QUERIES
Examiner	:	Michael N. Opsasnich
Group Art Unit :		2655

TRANSMITTAL LETTER

RECEIVED

SEP 17 2004

Technology Center 2600

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Transmitted herewith in triplicate is a <u>Supplemental Appeal Brief</u> to the Board of Patent Appeals, along with the following:

(X)   A Request for Reinstatement of Appeal; and
(X)   A return prepaid postcard.

If applicant has not requested a sufficient extension of time and/or has not paid any other fee in a sufficient amount to prevent the abandonment of this application, please consider this as a Request for an Extension for the required time period and/or authorization to charge our Deposit Account No. 11-1410 for any fee which may be due. Please credit any overpayment to Deposit Account No. 11-1410.

Ronald J. Schoenbaum
Registration No. 38,297
Attorney of Record
Customer No. 20,995
(949) 760-0404

H:\DOCS\ROS\ROS-6569.DOC\dns
091004

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Nicholas Lee	) Group Art Unit: 2655
			)
Appl. No.	:	09/650,173	)
			)
Filed	:	August 29, 2000	)
			)
For	:	VOICE INTERFACE AND	)
		METHODS FOR IMPROVING	)
		RECOGNITION ACCURACY	)
		OF VOICE SEARCH QUERIES	)
			)
Examiner	:	Opsasnick	)

RECEIVED

SEP 1 7 2004

Technology Center 2600

REQUEST FOR REINSTATEMENT OF APPEAL

Commissioner for Patents
Alexandria, VA 22313-1450

Dear Sir:

In response to the non-final Office Action mailed on August 11, 2004, Appellants hereby request reinstatement of the appeal pursuant to 37 C.F.R 1.193(b)(2)(ii). A Supplemental Appeal Brief is enclosed to address the new grounds for rejection. Please charge any fees that may be due, including any extension fees, to deposit account 11-1410.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: 9-10-04

By: _____
Ronald J. Schoenbaum
Reg. No. 38,297
Attorney of Record
Customer No. 20,995
949-721-2950

-1-

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Nicholas Lee	) Group Art Unit: 2655
			)
Appl. No.	:	09/650,173	)
			)
Filed	:	August 29, 2000	)
			)
For	:	VOICE INTERFACE AND	)
		METHODS FOR IMPROVING	)
		RECOGNITION ACCURACY	)
		OF VOICE SEARCH QUERIES	)
			)
Examiner	:	Opsasnick	)

<u>SUPPLEMENTAL APPEAL BRIEF</u>

RECEIVED

SEP 1 7 2004

Technology Center 2600

Board of Patent Appeals and Interferences
Washington, D.C. 20231

Dear Sir:

This Supplemental Appeal brief is being submitted in accordance with 37 C.F.R. 1.193(b)(2)(ii), together with a Request for Reinstatement of Appeal, to address the new grounds for rejection set forth in the non-final Office Action mailed on August 11, 2004.

The following sections of the original Appeal Brief filed on May 24, 2004 are hereby incorporated by reference:

III. Status of the Claims

IV. Status of Amendments

V. Summary of the Invention

VII. Grouping of Claims

VIII-A. Explanation of separate patentability of claim groups

Appendix (claim listing)

A. REAL PARTY IN INTEREST (UPDATED)

The real party in interest in the present application is now A9.com, Inc., which is a subsidiary of Amazon.com, Inc.

B. RELATED APPEALS OR INTERFERENCES (UPDATED)

An appeal is now pending in U.S. Appl. No. 09/729,646, filed December 4, 2000, titled GRAMMAR GENERATION FOR VOICE-BASED SEARCHES, which is owned by the assignee of the present application. An Appeal Brief has not yet been filed.

C. SUMMARY OF OFFICE ACTION

In the Office Action mailed on August 11, 2004 (hereinafter "the Office Action"), the Examiner rejected Claims 1-16, 20-26, 28, 29 and 31-55 on obviousness grounds over the combination of U.S. Patent No. 6,377,927 ("Loghmani et al"), U.S. Patent No. 5,917,889 ("Brotman"), and U.S. Patent No. 6,434,524 ("Weber '524").

The status of Claims 17-19 and 27 is not clear from the Office Action. Based on the last paragraph of page 7 of the Office Action, Appellants assume these claims are rejected on obviousness grounds over the aforementioned patents in combination with U.S. Patent No. 6,532,444 to Weber.

The status of Claim 30 is also unclear, as the Office Action does not set forth any basis for rejecting Claim 30. Appellant respectfully requests that the Examiner clarify the status of Claim 30 in his Reply Brief. Unless a basis for rejection is provided, Appellant requests that Claim 30 be treated as allowed, and be omitted from this appeal.

D. ISSUES PRESENTED ON APPEAL

In view of Appellant's election to group the dependent claims with their respective independent claims, the only issue presented on appeal is whether independent Claims 1, 15, 24, 33, 43 and 50 are properly rejected on obviousness grounds over the combination of Loghmani et al, Brotman, and Weber '524.

E. <u>DISCUSSION OF REFERENCES APPLIED BY EXAMINER</u>

In rejecting the independent claims, the Examiner relied solely on Loghmani et al, Brotman, and Weber '524, each of which is discussed below. For purposes of this appeal, Appellant will treat Loghmani et al and Weber '524 as prior art, but reserves the right to later disqualify one or both references as prior art.

1.   <u>Loghmani et al</u>

Loghmani et al discloses a voice-optimized database system that enables users to conduct database searches by voice. Each searchable item in the database is stored in association with an audio vector that characterizes the sound of a name or phrase associated with the item. The audio vector includes vector components having values for respective phonemes in the searchable item's name or phrase. See col. 4, lines 13-37. Multiple audio vectors may be stored for a given searchable item, each of which corresponds to a different phrase that may be uttered to search for the item.

To process a spoken query from a user, the spoken query is parsed based on the phonemes therein, and an audio vector is assigned to the spoken query. The assigned audio vector is then compared to the audio vectors associated with the searchable items in the database to search for items having an audio vector that is close to that of the search query. See col. 4, lines 38-55. The search results are then presented to the user. Thus, the spoken search query is processed without using a voice recognition grammar to initially convert the search query to text.

Loghmani et al also discloses that if the database does not support the use of audio vectors, an intermediate audio vector valuation module may be used to convert the phonemes in the spoken query to text, so that one or more textual versions of the query may be passed to the database. See column 8, lines 38-55.

2.   <u>Brotman</u>

Brotman discloses techniques for reliably capturing a string of characters specified by a user via a telephone. These techniques involve having the user both (1) utter all of the characters in the intended string, and (2) select the corresponding keys on the telephone keypad for each of these characters—either by depressing these keys or by uttering the number (0-9) of each such key. The character utterances and the keypad selections are then used in combination to predict the characters intended by the user, so that the likelihood of misrecognition events is reduced.

For example, to input the word "cat," the caller would utter the letters C-A-T, and would depress the corresponding telephone keys 2-2-8 (or alternatively, utter the numbers 2-2-8). The keypad selections would then be used to limit the possible interpretations of the character utterances. For example, to interpret the utterance of the letter "C," the system may treat "A," "B" and "C" (corresponding to the key for "2") as the only valid utterances. See column 3, lines 42-53. This is accomplished by using the telephone keypad entries to create a grammar that specifies the valid characters that may be uttered by the user. See column 4, lines 36-41.

Using this process, Brotman's system generates a string of characters that are predicted to be the characters intended by the user. This character string is then audibly output to the user, and the user is prompted to indicate, with a "yes" or "no" reply, whether the generated string is what the user intended. If the string has not been accurately captured, the user can continue to interact with the system until the intended string has been correctly identified. See Figure 2, blocks 670-730, and column 5, line 36 to column 6, line 10.

Brotman is not directed to the capture of search queries. Even if Brotman's method were used to capture search queries, it would not provide an efficient process for doing so. For example, a user wishing to submit the search query "Stephen King" would apparently have to utter all eleven letters of his name, and would also have to select the corresponding eleven keys on the telephone keypad. In contrast, in Appellant's preferred embodiment, the user could conduct this search by pressing the telephone keys containing the characters "S-T-E" (and/or uttering these characters) and uttering the name "Stephen King." The increase in efficiency over Brotman is even greater for longer queries.

Nothing can be taken from Loghmani et al that would improve this inefficiency in Brotman's character entry process. To the contrary, if the methods of Brotman and Loghmani et al were combined for purposes of capturing a user's search query, the user would additionally be burdened with having to utter the search query.

3.   Weber '524

Weber '524 discloses a user interface through which a user can interact with a computer system by voice. The user's utterances are interpreted in-part using context-specific voice recognition grammars that correspond to specific subjects such as "news," "weather," and "stocks." Col. 7, lines 14-30. Unlike the method disclosed in the present application, these context-specific grammars are not selected based on the user's entry of specific characters of the uttered term or phrase. Rather, they are apparently selected based on the topic or subject currently being browsed, as determined from prior utterances. See col. 3, lines 15-21 and col. 8, lines 45-53.

In contrast to the methods disclosed in the present application, Weber's method of using context-specific voice recognition grammars is not well suited for searching large domains of items, such as a domain of millions of book titles or music titles. If Weber's method were used for this purpose, the user would likely have to "drill down" through multiple levels of item categories and subcategories (e.g., books\fiction\mysteries); otherwise, the voice recognition grammars would most likely be too large to provide reliable voice recognition. In addition to being burdensome to users, such an approach would require the users to know how the items they are searching for are categorized. Brotman and Loghmani et al do not suggest a solution to this deficiency in Weber '524.

It is not clear from the Office Action whether, or to what extent, the Examiner is relying on Weber '524 in rejecting the independent claims of the present application. The portion of Weber '524 cited in connection with these claims is column 13, lines 13-24, which is directed to making user-specific updates to grammar files to accommodate the idiosyncrasies of individual users. None of the independent claims, however, require such a feature.

F.   DISCUSSION OF ISSUES ON APPEAL

The rejections of the claims in Groups 1-6 are improper because (1) the Examiner has not identified a motivation to combine Loghmani et al and Brotman, and (2) Loghmani et al, Brotman and Weber '524 do not disclose or suggest all of the limitations of any independent claim. Each of these two separate bases for reversing the obviousness rejections is discussed below.

1.   **The rejections of the claims of Groups 1-6 are improper because the Examiner has not identified a motivation to combine Brotman and Loghmani et al.**

In rejecting the claims of Groups 1-6, the Examiner has failed to identify a legally sufficient suggestion or motivation to combine Loghmani et al and Brotman. See M.P.E.P. § 2143.03. The only basis given by the Examiner for combining these two references is that the addition of Brotman's character entry process to Loghmani et al's voice-based search process would "reduce the domain of field choices, as well as improving the accuracy process in using the dual character input and follow-up speech verification" of Brotman. Office Action at page 3, first paragraph.

This asserted basis, however, fails to recognize that if Brotman's process were used to capture and verify a search query string intended by a user, there would be no need to additionally use Loghmani et al's process to interpret the user's utterance of the full search query. This is because Brotman's process allows the user to verify that the intended character string has been properly identified by the system. Once the search query string has been properly identified, it can be processed using conventional textual query processing methods, and Loghmani et al's spoken query processing methods become unnecessary.

Further, if both the Brotman process and the Loghmani et al process were used in combination to capture the intended search query, the user would apparently have to go through the unnecessarily burdensome process of (1) selecting the telephone keypad keys associated with all of the characters of the search query string, (2) uttering each character of the search query string, (3) verifying that the search string has been properly recognized by the system, and (4) uttering the search query. One skilled in the art would not be motivated to design a system that requires users to undergo such an unnecessarily burdensome process.

Because the Examiner has not identified a valid suggestion or motivation to combine Brotman and Loghmani et al, the obviousness rejections of the claims of Groups 1-6 are improper.

The Examiner's asserted basis for adding the teachings of Weber '524 to Brotman and Loghmani et al—namely to accommodate voice idiosyncrasies of individual users—does not appear to be pertinent to any of the independent claims.

2.   The rejections of the claims in Groups 1-6 are improper because Loghmani et al, Brotman and Weber '524 do not disclose or suggest all of the limitations of independent Claims 1, 15, 24, 33, 43 and 50, respectively.

The rejections of the claims in Groups 1-6 are also improper because, as set forth below, the independent claims include limitations that are not disclosed or suggested by the combination of Loghmani et al, Brotman and Weber '524 (hereinafter "the applied references"). See M.P.E.P. § 2143.03.

<u>Claim 1</u>

Claim 1 is directed to an embodiment in which the grammar used to interpret a voice query (i.e., an utterance of a search query) is a "dynamic grammar" that is generated after the user has submitted a set of characters that define a portion of the search query. The dynamic grammar is generated based at least in part on an identified subset of items that correspond to the set of characters received from the user. The claim reads as follows, with reference characters added for purposes of discussion:

> 1.   A method for improving voice recognition accuracy when a user submits a search query by voice to search a domain of items, the method comprising:
> (a) prompting a user to submit a set of characters of a voice query for searching the domain of items, and receiving the set of characters from the user, wherein the voice query is an utterance by the user of a search query, and the set of characters defines a portion of the search query;
> (b) in response to receiving the set of characters from the user, identifying a subset of items in the domain that correspond to the set of characters;
> (c) generating a dynamic grammar based at least in part on the subset of items, said grammar specifying valid utterances for interpreting the voice query;
> (d) prompting the user to submit the voice query, and receiving the voice query from the user; and
> (e) interpreting the voice query using the dynamic grammar.

The rejection of Claim 1 is improper because, inter alia, the applied references do not disclose or suggest prompting a user to submit a set of characters that "defines a portion of the search query," as set forth in subparagraph (a). In connection with these limitations, Appellant submits that the term "portion" has its ordinary meaning, which is "a section or quantity within a larger thing; a part of a whole." See dictionary.com, definition no. 1 for "portion."

In connection with this claim language, the Examiner relies on the character-string entry method disclosed in Brotman. Brotman, however, does not disclose a method in which the user is prompted to submit a set of characters "that defines a portion of" a search query or other intended character string. Rather, Brotman describes a character entry method in which the user both utters, and selects the corresponding telephone keypad keys for, <u>all</u> of the characters of the intended character string. Thus, and as explained above in the discussion of Brotman, if Brotman's character entry method were used to capture a search query, the user would apparently have to separately utter, and select the respective telephone key associated with, each character in the search query string. This approach would be highly burdensome to users, especially when submitting lengthy search queries.

The rejection is also improper because the applied references do not disclose or suggest the limitations of subparagraph (c), particularly when read in conjunction with subparagraph (b). In connection with these limitations, the Examiner relies on Figure 2, block 630 of Brotman. As explained at column 4, lines 36-41 of Brotman, block 630 of Figure 2 depicts a step in which each numerical telephone digit (0-9) selected by the user is used to limit the possible characters that can be validly uttered by the user. For instance, if the user depresses the keys "4," "7," "2," a grammar would be created that defines the allowable characters as {G,H,I}, {P,R,S}, and {A,B,C}. With this approach, the grammar is generated based solely on the numerical digits specified by the user. In contrast, Claim 1 recites a method in which a grammar is generated based at least in part on a subset of items, within a domain of items being searched, that correspond to a set of characters received from the user.

Because Claim 1 includes limitations that are not disclosed or suggested by the applied references, the obviousness rejection of the claims of Group 1 (Claims 1-14, 39 and 40) is improper.

Claim 15

Claim 15 is directed to a method for improving voice recognition accuracy when a user submits a query by voice to search a domain of items. The method comprises "receiving a set of characters entered by a user, the set of characters representing a portion of a query." The applied references do not disclose or suggest this step. As mentioned above, if Brotman's character entry

method were used to capture a query, the received characters would be those of the entire query, and not a portion of the query.

Claim 15 also recites "in response to receiving the set of characters, selecting a grammar which is derived at least in-part from text extracted from a subset of items that correspond to the set of characters entered by the user;" and "providing the grammar to a voice recognition system for use in interpreting the query as entered by the user by voice." (Note that the "whereby" clause in Claim 15 clarifies that the phrase "the query as entered by the user by voice" refers to an utterance by the user of the full query.) The applied references do not disclose or suggest these limitations. The Examiner did not fully address these limitations in the Office Action.

The applied references also fail to disclose or suggest a method in which "the user's entry of a subset of characters of the query, together with the user's utterance of the full query, are used in combination to capture the query." The Examiner did not fully address these limitations (and particularly the phrase "a subset of characters of the query") in the Office Action.

Because the applied references fail to disclose or suggest all of the limitation of Claim 15, the obviousness rejection of the claims of Group 2 (Claims 15-23) is improper.

<div align="center">Claim 24</div>

Claim 24 is directed to a system that includes "a first code module which causes a user to be prompted to enter a set of characters of a query such that the user may partially specify the query," and "a second code module which causes the user to be prompted to utter the query." The claim also calls for a query server that "is programmed to use the set of characters to select a grammar for use by [a] voice recognition system to interpret the query as uttered by the user."

The applied references do not disclose or suggest such a system. In this regard, if Brotman's character entry method were used to capture search queries, the user would not enter a set of characters that partially specifies the query, but rather would individually specify all of the characters of the search query.

Further, even if Brotman's method were used to capture a set of characters of the query, it would not be obvious from the applied references to then use this captured set of characters to select a grammar for interpreting the user's utterance of the query. The Examiner has not identified any disclosure in the applied references that suggests this aspect of the claimed method.

Because Claim 24 includes limitations that are not disclosed or suggested by the applied references, the obviousness rejection of the claims of Group 3 (Claims 24-32) is improper.

Claim 33

Claim 33 involves a method in which a user refines a search query by uttering an additional query term to add to the query. To interpret the user's utterance of the additional query term, a grammar is generated, at least in-part, by extracting text from the set of search result items resulting from the query. A preferred embodiment of this method is described beginning at page 9, line 24 of the present application.

The Examiner's analysis of Claim 33 at the top of page 9 of the Office Action fails to address many of the limitations of this claim. For example, the Examiner appears to disregard the limitations "providing the user an option to refine the query by adding an additional query term," "generating a grammar at least in-part by extracting text from the set of search result items," and "using the grammar to interpret an utterance by the user of an additional query term." These limitations are not disclosed or suggested by the applied references. Indeed, none of the applied references discloses a query refinement process, let alone the particular query refinement process defined in Claim 33.

Because Claim 33 includes limitations that are not disclosed or suggested by the applied references, the obviousness rejection of the claims of Group 4 (Claims 33-38) is improper.

Claim 43

Claim 43 is directed to a method that involves "prompting a user to depress a sequence of telephone keypad keys corresponding to a sequence of characters of a query term of a search query." The user is also prompted "to utter the search query by voice." The voice utterance of the search query is interpreted "using a voice recognition grammar that corresponds to the sequence of keys depressed by the user."

Of the applied references, Brotman is the only reference that involves the use of a voice recognition grammar that is based on a sequence of telephone keypad keys depressed by a user. Brotman's grammar, however, is suitable only for interpreting utterances of individual characters, and not for interpreting an utterance of a search query. Thus, even if Brotman's teaching to create a grammar based on a user's telephone keypad entries were combined with the teachings of Loghmani et al and/or Weber '524, the combination would not involve "interpreting

the voice utterance [of the search query] using a voice recognition grammar that corresponds to the sequence of keys depressed by the user," as required by Claim 43.

Because Claim 43 includes limitations that are not disclosed or suggested by the applied references, the obviousness rejection of the claims of Group 5 (Claims 43-47) is improper.

<u>Claim 50</u>

Claim 50 is directed to a method of capturing a search query specified by a user by telephone. The method comprises "receiving from the user an indication of a subset of the characters contained in the search query, said indication of the subset of characters being specified at least in part as telephone keypad entries." The method further comprises "receiving from the user a voice utterance that represents the entire search query," and "interpreting the voice utterance using a voice recognition grammar that corresponds to the indication of the subset of characters."

The rejection of Claim 50 is improper because, inter alia, the applied references do not disclose or suggest "interpreting the voice utterance [that represents the entire search query] using a voice recognition grammar that corresponds to the indication of the subset of characters [contained in the search query]." As discussed in connection with Claim 43, Brotman is the only applied reference that involves the use of a voice recognition grammar that is based on telephone keypad entries, and this grammar would not be suitable for interpreting a voice utterance of a search query. Further, Brotman's grammar is created based on the keypad entries corresponding to the entire character string to be captured, and not a "subset of characters" as claimed.

Because Claim 50 includes limitations that are not disclosed or suggested by the applied references, the obviousness rejection of the claims of Group 6 (Claims 50-55) is improper.

G. <u>CONCLUSION</u>

For the reasons set forth above, Appellant submits that the rejections of the claims of Groups 1-6 are improper, and requests that these rejections be reversed.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: 9-10-04   By: _____

Ronald J. Schoenbaum
Reg. No. 38,297
Attorney of Record
2040 Main Street, 14th Floor
Irvine, CA, 92614
Customer No. 20,995
949-721-2950